

Association for Information Systems

## AIS Electronic Library (AISeL)

---

Proceedings of the 2018 AIS SIGED  
International Conference on Information  
Systems Education and Research

SIGED: IAIM Conference

---

2018

# A QUANTITATIVE CONSIDERATION OF RIGOR AND RETALIATION: INVESTIGATING STUDENT EVALUATIONS OF TEACHING, RIGOR, AND GRADE DISTRIBUTIONS

Jeremy Shafer

Department of Management Information Systems Temple University, [jeremy.shafer@temple.edu](mailto:jeremy.shafer@temple.edu)

Follow this and additional works at: <https://aisel.aisnet.org/siged2018>

---

### Recommended Citation

Shafer, Jeremy, "A QUANTITATIVE CONSIDERATION OF RIGOR AND RETALIATION: INVESTIGATING STUDENT EVALUATIONS OF TEACHING, RIGOR, AND GRADE DISTRIBUTIONS" (2018). *Proceedings of the 2018 AIS SIGED International Conference on Information Systems Education and Research*. 23.  
<https://aisel.aisnet.org/siged2018/23>

This material is brought to you by the SIGED: IAIM Conference at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2018 AIS SIGED International Conference on Information Systems Education and Research by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

## **A QUANTITATIVE CONSIDERATION OF RIGOR AND RETALIATION: INVESTIGATING STUDENT EVALUATIONS OF TEACHING, RIGOR, AND GRADE DISTRIBUTIONS**

Jeremy Shafer  
Department of Management Information Systems  
Temple University  
Jeremy.Shafer@temple.edu

### **Abstract:**

This paper introduces a numeric measure of course rigor as it pertains to grade distributions. The nature of this key performance indicator (KPI) is described in detail, along with statistical tests indicating that instructors of classes with rigorous grade distributions need not necessarily fear retaliation in the form of low student evaluation scores.

The paper also describes ways in which the administrators have used the KPI to reinforce expectations for rigor in grading practices and to recognize rigorous instructors. The KPI and practices related to it are offered as counter weight to forces which might encourage lenient grading practices.

**Keywords:** rigor, grades, grade inflation, KPI, education, administration, student evaluations of teaching (SET)

### **I. INTRODUCTION**

An instructor of an undergraduate college course rarely intends to grade leniently. Similarly, instructors rarely intend to deliberately create low rigor, “easy A” course material. However, there are incentives that entice instructors to do so [Millet, 2018]. Instructors often believe that lenient grading and/or lighter material will result in increased student satisfaction as expressed in formal student evaluations of teaching (SET) scores [Jewell and McPherson, 2012]. Similarly, it is conceivable on a priori grounds that instructors may resist injunctions from their administration to grade rigorously for fear that poorer performing students will ultimately punish them with critical feedback and low teacher evaluation scores.

In 2015, the Management Information Systems (MIS) Department at Temple University created a new unique KPI based on the ratio of “high” grades to “low” grades in each course. This KPI was termed the Grade Rigor Ratio (GRR). It affords the department several distinct benefits. First, over time the GRR KPI became a proven and convenient mechanism for the department to initiate discussions with individual instructors about rigor. Secondly, course coordinators use the GRR KPI to communicate grading rigor expectations to fellow instructors. Third, the GRR KPI is a catalyst for individual instructors to reflect on their curriculum development, instruction and grading practices. Fourth, achieving departmental GRR expectations has now become a basis for recognizing faculty excellence. This provides for the first time, administrators a tool other than student evaluations for rewarding and recognizing faculty.

This paper provides a quantitative analysis of the GRR. The results can form the basis to discuss the objections of instructors concerning departmental directives to increase rigor as well as the notion that students will punish demanding instructors by assigning them low course evaluation scores. If there is no clear dependency between grading rigor (as measured by GRR) and student feedback scores (as collected in the SET), then instructors might become emboldened to let the grades they assign reflect a truer assessment of each student’s ability or lack thereof.

This paper describes the GRR KPI, its introduction to the Temple University MIS Department and it summarizes the findings. It also statistically considers the question of how SET data may vary in response to rigor as measured by GRR. This paper also reflects on the GRR KPI’s impact on

instructors as they plan and execute their courses. Possibilities for further research are also recommended.

## II. LITERATURE REVIEW

Grades are a signal of quality to “parents, employers, postsecondary gatekeepers, and students themselves” [Pattison et al., 2013]. Therefore, it follows that lenient grading, grade inflation, or (worst of all) low-rigor instruction would weaken the value of that signal.

The problem of grade inflation has been an issue for decades and continues to be an issue in American universities [Jaschik, 2016]. Similarly, the problems of grading leniency and grade inflation are cause for concern in every discipline [Germain and Scandura, 2005]. Some disciplines appear to be more susceptible to grade inflation. The literature states that ‘Economics, Chemistry, and Math are consistently low-grading departments, while Art, English, Music, Philosophy, Psychology, and Political Science are almost always high-grading departments’ [Sabot and Wakeman-Linn, 1991].

Previous research has explored the relationship between high grades and favorable student evaluations. The empirical evidence for a direct relationship between grades and SET performance is mixed [Isely and Singh, 2005].

The usefulness of SET data itself has been challenged by some educators. Its use by administration for faculty re-appointment decisions, promotion, and tenure has been the source of some concern, with some educators growing hostile to SET evaluations in general [Germain and Scandura, 2005].

Meanwhile, educators in MIS and related disciplines continue to explore means of encouraging genuine learning [Serva and Monk, 2014] and fostering a sense of self-efficacy [Smith et al., 2003] as well as team-working skills [Bond et al., 2015]. SET data and similar survey instruments are used in the evaluation of these efforts. It follows that, while SET instruments may have deficiencies, they are commonly used in the absence of any superior quantitative instrument.

In summary, this paper does not seek to challenge or defend the practice of collecting and referencing SET data. Instead, the paper proposes a complementary tool, the GRR, to assist faculty and administrators who are interested in improving rigor.

## III. THE GRR OBJECTIVE

The GRR KPI was born out of a desire for a quantitative measure of rigor in grading practices. The basic premise behind the GRR KPI is that demanding content, delivered to a diversely abled population of students should result in varied, stratified results, along with equally varied and stratified grades.

The GRR KPI itself is a measure of the final grade distribution of the course. It is not, in itself, a measure of rigorous instruction. It is, however, indicative of an instructor’s willingness to grade rigorously. There are mechanisms in place at most universities that would discourage an instructor from making capricious grading decisions in order to adhere to the GRR KPI: the SET process, peer review of teaching, and a formal student grievance procedure. Consequently, while the GRR KPI is not a direct measure of rigorous instruction, it can be thought of as a surrogate for such rigor.

The Temple MIS department created the GRR KPI with the intent to direct instructor behavior in the following ways:

1. Instructors should clearly identify students who are succeeding in the classroom versus those students who are not. Great students need to know that they are accomplishing

something significant in their studies and that their high performance will ultimately set them apart from their peers in both their chosen field and in other areas of study. Conversely, poor performing students need to know that they are performing below the expectations of a competitive program. While never pleasant, the difficult conversations that go with such an operative philosophy can (potentially) help students rise to the expectations of their chosen profession or seek an alternate career trajectory that is a better match for their strengths and preferences.

2. Instructors should think critically about their grades they award, and the probable outcomes for each course given their knowledge about the student population. For example, if a particular topic (say, programming) is known to be difficult, and there is awareness that not all students apply themselves equally to learning, and that not all students enter the class with the same level of readiness and natural ability, then the resulting class grades should reflect that diversity. The instructor that endorses every student with a good or excellent grade is not accurately representing reality.

#### IV. THE DEFINITION OF GRR

The Grade Rigor Ratio is defined as follows: the GRR KPI is the ratio of the percentage of “high” grades (B+ or better) over the percentage of “low” grades (C- or below) in a single section of a course. The percentage of “low” grades also includes the number of students who withdrew from the course after the university add/drop date.

Examples:

- A GRR of 2.0 indicates that the instructor awarded twice as many “high” grades as “low” grades.
- A GRR of 1.0 indicates that the ratio of “high” to “low” grades is perfectly balanced. The instructor awarded an equal number of “high” and “low” grades.

The value of .001 is added to both the numerator and denominator of the GRR calculation. Adding this value avoids the inevitable divide-by-zero error that occurs when no “low” grades are awarded by the instructor at all.

The GRR calculation also includes a conditional statement. If the percentage of medium grades is more than 68%, the formula will report the calculated GRR as well as the text “POOR SPREAD”. The “POOR SPREAD” text indicates that instructor is not effectively differentiating high performing students from low performing ones. That is, if more than 68% of the students in a class receive medium range grades, then this is an unusual concentration of medium grades. This concentration in the medium range is contrary to the departmental objectives that motivated the creation of GRR KPI.

It should be noted that, if GRR were used in another institutional context, the definitions of “high” and “low” will be different. In the Temple University MIS department, a grade of C or better is required to move forward in the MIS program of study. Consequently, a grade of C- is truly a “low” grade for a Temple MIS student and it carries consequences similar to grades of D or F as may be defined in other departments and institutions. MIS students are allowed two attempts to earn of grade of C or better in an MIS course before being removed from the program.

#### V. THE INTRODUCTION OF GRR

The introduction of GRR within the department was not without controversy. One initial misconception was that the use of GRR would force instructors to meet a completely arbitrary quota of “high” and “low” grades. The GRR KPI is defined in such a way that it allows instructors latitude

in how they distribute grades in their classes. The following diagrams are similar to ones used in conversations with MIS faculty when the GRR KPI was introduced.

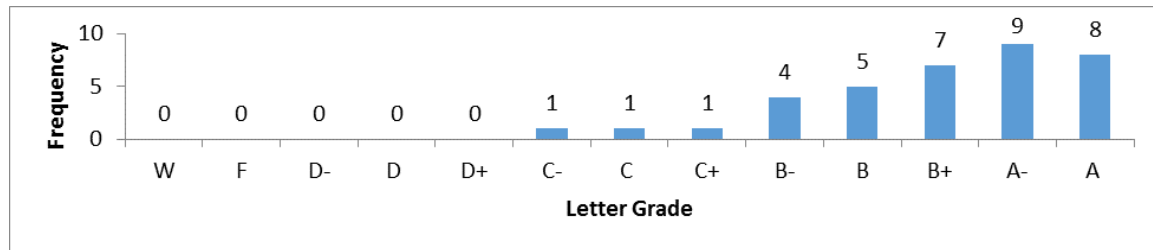


Figure 1: Too lenient

In the hypothetical class of 36 students shown in Figure 1, the  $GRR = (8 + 9 + 7) / (1) = 24$ . Given the diverse levels of readiness and commitment in the student population, a GRR KPI value of 24 would be unacceptable for most undergraduate courses.

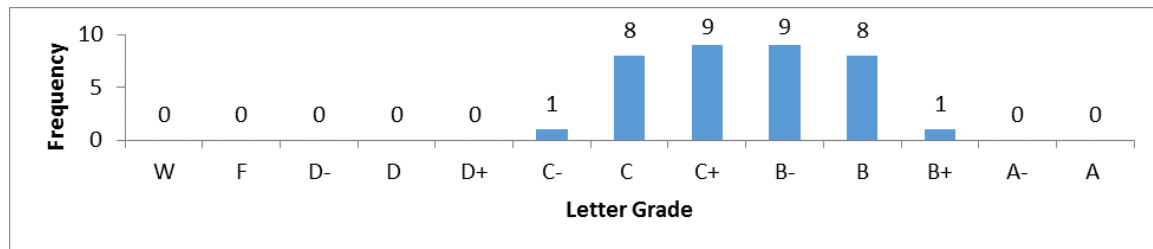


Figure 2: Poor Spread

In the hypothetical class shown in Figure 2, the  $GRR = (1) / (1) = 1.0$  POOR SPREAD. According to the GRR KPI, no more than 68% of the letter grades should be between C and B, inclusive. This GRR KPI reports "POOR SPREAD" for this data because 94% of the grades in this hypothetical class are neither high nor low in Figure 2. Grading that avoids extremes does not provide a useful signal of student quality/comprehension. For this reason, the GRR KPI for Figure 2 is unacceptable.

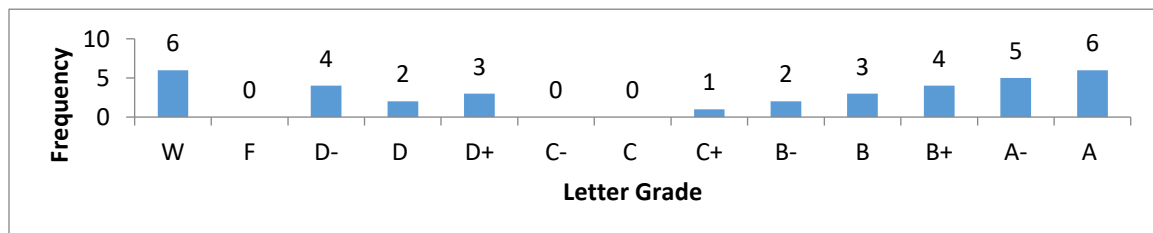


Figure 3: Bimodal

In the class shown in Figure 3 the  $GRR = (6 + 5 + 4) / (3 + 2 + 4 + 6) = 15 / 15 = 1$ . This is a strong GRR score. This grade distribution suggests that the instructor is willing to award grades that identify which students have met the learning objectives of the course, and which students have not. Notice that this distribution does not follow a normal distribution. That is perfectly acceptable according to the GRR KPI expectations of the department.

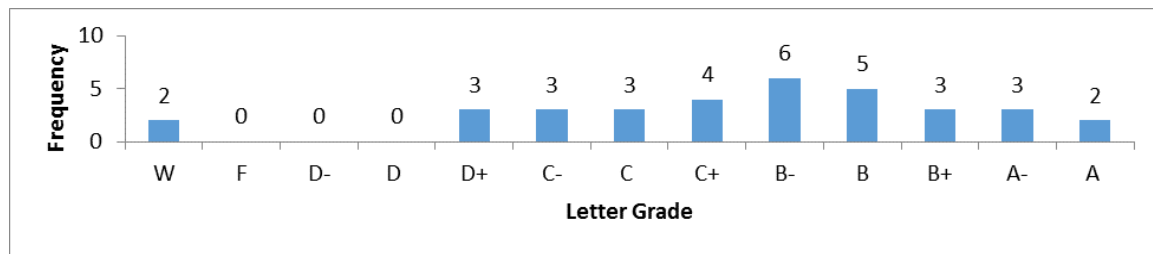


Figure 4: Normal

In the hypothetical class of 36 students in Figure 4, the  $GRR = (2 + 3 + 3) / (3 + 3 + 2) = 8 / 8 = 1$ . This is a strong (good) GRR score. The grade distribution in figure 4 is closer to a normal distribution. This too is acceptable to the department.

While it is true that students (and human beings in general) can be inconsistent, most instructors found that they could exert great influence over the GRR outcome in their classes by carefully considering the difficulty of instructional content, exam questions and the weights given to the various items used to calculate the class grade. The key point is that since the GRR KPI is a ratio it avoids the trap of requiring a particular distribution, or the need to award a certain number of grades.

Another objection was that the GRR KPI would be used in a “one-size-fits-all” manner. This concern diminished as several semesters of GRR data were compiled, and varying thresholds of expectation were established and adjusted over time. For example: it has become a departmental norm that any GRR KPI at or below a 2.0 is considered satisfactory for an undergraduate course in our core-curriculum. However, an MIS General Education course offered to incoming freshmen is expected to have a GRR between 2.0 and 9.0, inclusive. Other undergraduate courses are expected to have a GRR of 5.0 or less.

The above demonstrates another benefit of the GRR KPI. It can be used to more precisely communicate the expectations between faculty and departmental administration. For example: a GRR of 4.0 indicates that an instructor awarded four times as many high grades as they did low grades in a particular class. With the GRR KPI the department can clearly articulate that this is an acceptable outcome for an undergraduate elective course but not a core course. The resulting departmental culture then becomes very different from one with no articulated expectations regarding rigor at all.

All measures require some kind of boundary, i.e., conditions where they are not applicable. So it was decided that honors sections of courses, courses with fewer than 15 students, and sections of our capstone course are exempt from GRR review. In the case of the capstone, the rationale was that since students are graduating there is a tendency for grades to converge to a higher level and the department was comfortable with this scenario.

To summarize, the major ongoing concern of the faculty remained the following: students will retaliate against strong GRR instructors by awarding them poor SET scores. This is a reasonable concern because SET scores are used for merit, promotion and tenure decisions in most universities.

After repeated assurances from the department as well as the promise of recognition for rigor, many instructors chose to introduce more rigorous content and assessments. They did this because there was a near universal understanding of the long-term benefits of rigor, and because of various GRR related incentives the department's administration introduced. The acknowledgment and/or reward of instructors that held their students to high standards was an important factor leading to the acceptance of the GRR KPI. These incentives are described later in this document.

## VI. GRR AND RETALIATION

This section investigates the relationship between GRR and SET. The SET instrument collects responses to many questions that are of interest. But of primary interest here are the responses to three SET questions that would have the most immediate bearing on the academic career of an instructor. If students can be shown as awarding lower scores on these questions to instructors of noticeably strong GRR courses, then the students can be thought of as engaging in a form of retaliation against those instructors

Undergraduate MIS course SET data from two full academic years was compiled. The compiled data set contained 163 records, with each record representing one section of an undergraduate course.

As previously mentioned, some courses (honors sections, courses contain fewer than 15 students and the capstone course) are exempt from GRR review. These were removed from the data set. Similarly, 8 courses with "POOR SPREAD" indicators were removed, as they were not clearly indicative of an instructor seeking to delineate high performing students from low performing students. After this, 131 records remained in the data set.

In order to investigate the impact of deliberate efforts to raise rigor (both in grading and in coursework) the strongest GRR courses were contrasted with the rest of the MIS undergraduate courses. The strongest GRR category defined by the department was  $GRR \leq 2.0$ .

It is useful to look at the aggregated grade distributions of these strong GRR courses in comparison to their inconspicuous counterparts.

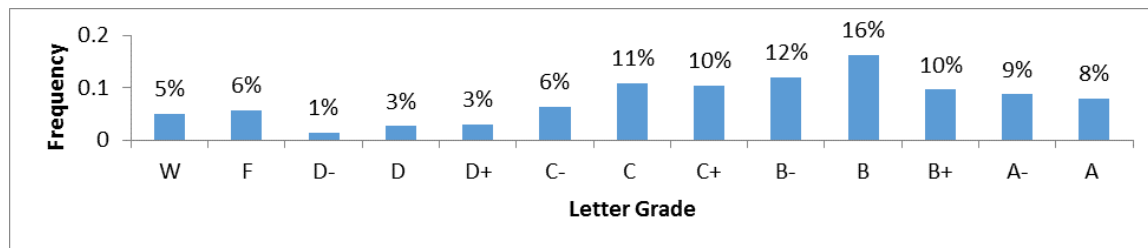


Figure 5: Strong GRR

Figure 5, shown above, represents the aggregated distribution of grades given in courses with a  $GRR \leq 2.0$ . There were 71 such courses in the timeframe. These are singled out because achieving a GRR KPI  $\leq 2.0$  requires that the instructor deliver demanding content and grade judiciously.

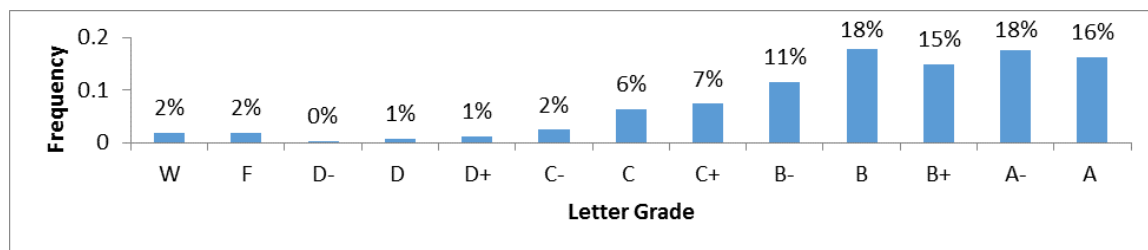


Figure 6: Inconspicuous GRR

Figure 6, shown above, represents the aggregated distribution of grades given in courses with a  $GRR > 2.0$ . For the purposes of this document, GRR KPI scores greater than 2.0 will be referred to as Inconspicuous GRR scores. There were 60 such courses in the timeframe.

With these two groups (strong GRR and inconspicuous GRR) the following hypothesis are of interest.

## Hypothesis 1 - Instructor Quality

*H1: The average SET scores received by strong GRR course instructors are different from the average SET scores received by the instructors of inconspicuous GRR courses. (That is, students retaliate against strong GRR course instructors.)*

In the SET data the question “The instructor taught this course well” is considered a prominent indicator of instructor quality. The students respond to this question on a Likert-type scale that ranges from 1 to 5.

On average, courses with an inconspicuous GRR received a higher SET score ( $M = 4.26$ ,  $SE = .05$ ) than courses with a strong GRR ( $M = 4.17$ ,  $SE .04$ ). This difference, .081 [CI -.06, .23] was not significant,  $t(121.5) = 1.07$ ,  $p = 0.05$ , with an effect size of  $d = -.20$ .

Overall, the results do not show evidence to accept H1.

## Hypothesis 2 - Student Learning

*H2: Students will report learning more in strong GRR courses.*

In the SET data the question “I learned a great deal in this course” is commonly used to gauge student’s sense of learning. The students respond to this question on a Likert-type scale that ranges from 1 to 5.

On average, courses with an inconspicuous GRR received a higher SET score ( $M = 4.17$ ,  $SE = .052$ ) than courses with a strong GRR ( $M = 4.14$ ,  $SE .06$ ). This difference, .02 [CI -.13, .17] was not significant,  $t(128.8) = .281$ ,  $p = 0.05$ , with an effect size of  $d = -.05$ .

Overall, the results do not show evidence to accept H2.

## Hypothesis 3 - Student Time Investment

*H3: Students will report spending more time in strong GRR courses.*

In the SET data the question “On average, hours per week spent preparing for class and completing course assignments” is commonly used to gauge student’s time commitment to a class. The students respond to this question on a Likert-type scale that ranges from 1 to 7. A score of 7 translates to “I spent 8 or more hours a week outside of class”, while a score of 1 translates to “I spent less than 1 hour a week”.

On average, courses with an inconspicuous GRR received a lower SET score ( $M = 3.65$ ,  $SE = .087$ ) than courses with a strong GRR ( $M = 4.31$ ,  $SE .078$ ). This difference, -.65 [CI -.89, -.42] was significant,  $t(124.5) = -5.58$ ,  $p = 0.05$ , with an effect size of  $d = .97$  which is a large size effect.

Overall, the results show evidence to accept H3.

## VII. DISCUSSION

The results do not show evidence of a positive or negative relationship between SET scores and GRR. In other words, contrary to popular expectations, there is no evidence that students will retaliate against instructors for offering more rigorous courses than other instructors’. Further, in comparison there is also no evidence that offering less rigorous courses will lead to improved SET scores. Although a lack of evidence cannot prove causality, the results are still interesting, especially since high GRR does positively influence the average time spent.

These findings may come as a surprise to both instructors and their administration. The findings suggest that instructors can hold their students to high standards without fear of retaliation. The findings may also suggest that instructors who do not receive satisfactory student feedback may need to more critically consider the quality of their instruction and related materials instead of gravitating towards solutions that attempt to garner favor with students by grading more leniently.



One explanation for the finding related to H1 is that poor performing students will often withdraw from a course. Consequently, even if those students do harbor resentment towards the instructor, they will not be present for the SET survey at the end of the semester. Therefore, it is possible for instructors to maintain satisfactory (or even excellent) SET scores while simultaneously holding their students to high standards. It should also be noted that strong GRR ( $GRR \leq 2.0$ ) does not appear to have a relationship with learning as reported by the students (H2). However, the data does suggest a dependency between strong GRR and time spent as reported by the student (H3). This may be an opportunity for further study. It may be that the students who invest more time in the strong GRR courses do not immediately recognize the benefit of their time investment. A longitudinal study of students in strong GRR versus weak GRR courses may shed further light on this. A thematic analysis of the unstructured text collected in SET for those courses which have both strong GRR and excellent SET scores would also be of interest.

When confronted with a genuinely rigorous, demanding course some students will choose to withdraw in order to preserve an otherwise satisfactory GPA and subsequently return to the class at a later date better equipped to succeed. Other students will find that the class helps them understand the demands of an MIS / IT career and they consequently choose a new major better suited to their strengths.

As noted earlier, the Temple MIS department administration promoted the adoption of the GRR KPI in a number of ways, including: Instructors who do not routinely meet GRR expectations may become ineligible for teaching awards and/or merit adjustments related to teaching. Similarly, instructors who do meet GRR expectations receive preferential treatment in class scheduling. Further development of such incentive mechanisms, and their varying utility, could be a topic for further study. Clearly, additional study is needed to assess if the GRR KPI is relevant and useful in other fields, universities, and programs.

The presence of the GRR KPI, its visibility, and the incentives and disincentives that are tied to it by the department encourage instructors to think critically about rigor at each stage of a course's development. When the syllabus is being drafted, the weights of exams, quizzes, assignments and projects are constructed with more caution. During the semester, grading will (out of necessity) be tough, but fair. As the semester draws to a close, grade adjustments and/or other forms of special consideration that students often request will be viewed with greater scrutiny.

## CONCLUSION

The GRR KPI is offered here for use in our academic programs; a critical value-add within the MIS community. Educators in the field of MIS have the opportunity (and the responsibility) to manage the student population that will ultimately define the industry. If this population is to be thoughtfully managed, then a KPI such as GRR is a useful instrument for such an effort.

## ACKNOWLEDGEMENTS

The author wishes to thank the past and present chairs of the MIS Department at Temple University: Munir Mandviwalla and David Schuff. The author also wishes to thank Professor Mart Doyle for his mentorship. All of the above were instrumental to creation and application of the GRR KPI. The author appreciates their ongoing support.

Bond, J., Raju, P. K., and Sankar, C. (2015). Influence of Team Skills in Engineering Technology. Retrieved from <https://aisel.aisnet.org/amcis2015/ISEdu/GeneralPresentations/28/>

Germain, M. L., and Scandura, T. A. (2005). Grade Inflation And Student Individual Differences as Systematic Bias in Faculty Evaluations. *Journal of Instructional Psychology*, 32(1). Retrieved from [http://people.uncw.edu/caropresoe/EDN523/523\\_Spr\\_07/Grade\\_Inflation.pdf](http://people.uncw.edu/caropresoe/EDN523/523_Spr_07/Grade_Inflation.pdf)

Isely, Paul and Singh, Harinder (2005) "Do Higher Grades Lead to Favorable Student Evaluations?". *The Journal of Economic Education*, 36:1, 29-42, DOI: 10.3200/JECE.36.1.29-

42. Retrieved from  
<https://doi.org/10.3200/JECE.36.1.29-42>
- Jaschik, S. (2016, March 29). Grade Inflation, Higher and Higher. Retrieved from  
<https://www.insidehighered.com/news/2016/03/29/survey-finds-grade-inflation-continues-rise-four-year-colleges-not-community-college>
- Jewell, R., and McPherson, M. (2012). Instructor-Specific Grade Inflation: Incentives, Gender, and Ethnicity. *Social Science Quarterly*, 93(1), 95-109. Retrieved from  
<http://www.jstor.org/stable/42864061>
- Millet, I. (2018). The relationship between grading leniency and grading reliability. *Studies in Higher Education*, 43(9), 1524-1535. Retrieved from  
<https://www.tandfonline.com/doi/full/10.1080/03075079.2016.1265497>
- Pattison, E., Grodsky, E., and Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259-265. Retrieved from  
<http://journals.sagepub.com/doi/abs/10.3102/0013189X13481382>
- Sabot, R., and Wakeman-Linn, J. (1991). Grade inflation and course choice. *Journal of Economic Perspectives*, 5(1), 159-170. Retrieved from  
<https://www.aeaweb.org/articles?id=10.1257/jep.5.1.159>
- Serva, Mark A. and Monk, Ellen (2014) "BEYOND THE FIVE-MINUTE UNIVERSITY: IMPROVING LEARNING IN OUR MIS CLASSES". 2014 Proceedings. 11. Retrieved from  
<https://aisel.aisnet.org/siged2014/11>
- Smith, Deborah; Chang, Jerry; and Moores, Trevor (2003) "Comparing Self-Efficacy and Meta-Cognition as Indicators of Learning". AMCIS 2003 Proceedings. 409. Retrieved from  
<https://aisel.aisnet.org/amcis2003/409>